

Ανάπτυξη Εφαρμογής για Ανάλυση Δεδομένων

Ομαδική Αναφορά

September 15, 2024

1 Εισαγωγή

Η δημιουργία εφαρμογών για ανάλυση δεδομένων είναι ουσιαστική για την απόκτηση πολύτιμων πληροφοριών από τα δεδομένα. Αυτή η εφαρμογή έχει αναπτυχθεί για να προσφέρει εύχρηστα εργαλεία για τη φόρτωση, επεξεργασία και οπτικοποίηση δεδομένων, καθώς και για την εφαρμογή αλγορίθμων μηχανικής μάθησης.

2 Σχεδιασμός της Εφαρμογής

Η εφαρμογή αναπτύχθηκε με τη χρήση της αρχιτεκτονικής MVC (Model-View-Controller). Το παρακάτω UML διάγραμμα απεικονίζει τη δομή της εφαρμογής.

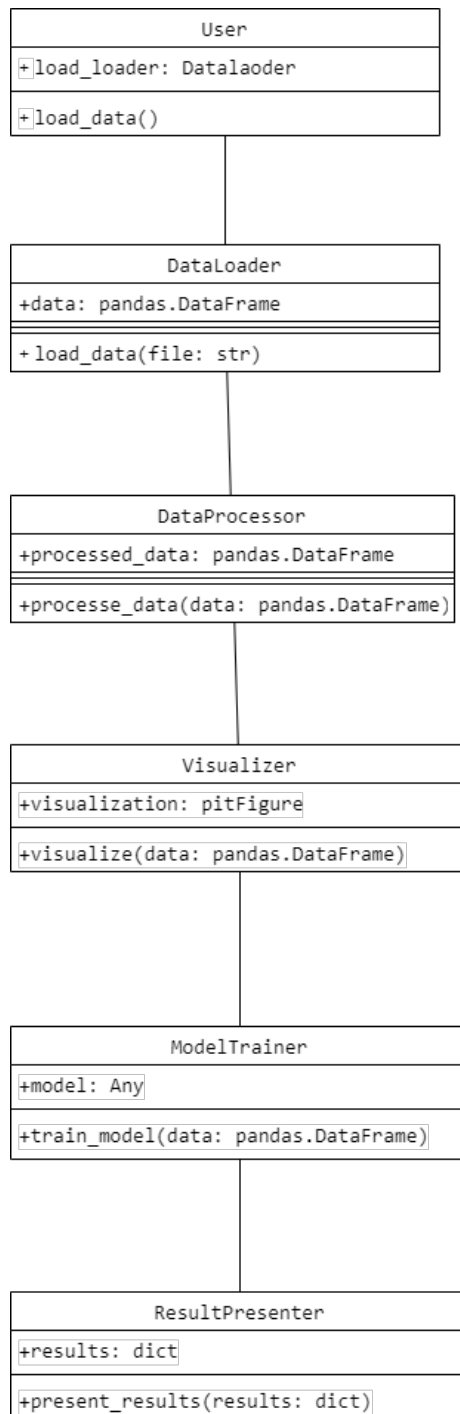


Figure 1: UML Διάγραμμα της Εφαρμογής

3 Υλοποίηση

Η υλοποίηση της εφαρμογής έγινε με τη χρήση της γλώσσας προγραμματισμού Python και των βιβλιοθηκών Streamlit, scikit-learn και matplotlib. Παρακάτω παρατίθεται ένα δείγμα κώδικα για τη φόρτωση και την επεξεργασία των δεδομένων.

```
import streamlit as st
import pandas as pd

# Φόρτωση δεδομένων
uploaded_file = st.file_uploader("Επιλέξτε ένα αρχείο CSV ή Excel", type=["csv", "xlsx"])
if uploaded_file is not None:
    data = pd.read_csv(uploaded_file) if uploaded_file.name.endswith('.csv') else pd.read_excel(uploaded_file)
    st.write("Data Preview:")
    st.write(data.head())
```

4 Αποτελέσματα

Η εφαρμογή διαθέτει εργαλεία για την ανάλυση δεδομένων με 2D οπτικοποιήσεις και χρήση αλγορίθμων μηχανικής μάθησης. Τα αποτελέσματα αυτών των αλγορίθμων παρουσιάζονται αναλυτικά, περιλαμβάνοντας τη σύγκριση των αποδόσεών τους.

5 Συμπεράσματα

Η εφαρμογή αποδεικνύει την αξία των εργαλείων ανάλυσης δεδομένων για την εξαγωγή χρήσιμων πληροφοριών. Η αξιοποίηση Python και Streamlit προσφέρει την ανάπτυξη ευέλικτων και επεκτάσιμων λύσεων.

6 Ομάδα Ανάπτυξης

Η ανάπτυξη της εφαρμογής πραγματοποιήθηκε από την ομάδα μας στο πλαίσιο του μαθήματος Software Engineering. Κάθε μέλος της ομάδας είχε συγκεκριμένες αρμοδιότητες, όπως περιγράφεται παρακάτω:

- **Θεοδόσης Καραγεωργίου (Inf2021076):** Ανάπτυξη Κώδικα, Υλοποίηση και Τεκμηρίωση
- **Κωνσταντίνος Λιάβας (Inf2021121):** Ανάλυση Δεδομένων, Προετοιμασία και Καθαρισμός Δεδομένων
- **Φοίβος Ελευθερίου (Inf2021049):** Οπτικοποίηση Δεδομένων, Δημιουργία Διαγραμμάτων και Γραφημάτων
- **Όλοι μαζί:** Συνεργασία στον Σχεδιασμό και Δοκιμές

7 Κύκλος Ζωής Έκδοσης Λογισμικού

Για την ανάπτυξη της εφαρμογής, επιλέξαμε το μοντέλο Agile, καθώς αυτό είναι ένα προσαρμοστικό μοντέλο που επιτρέπει τη συνεχή βελτίωση του προϊόντος μέσα από επαναληπτικούς κύκλους. Ακολουθεί μια σύντομη εξήγηση του τρόπου που εφαρμόσαμε το Agile στη διαδικασία ανάπτυξης:

7.1 Στάδια του Agile

1. **Σχεδιασμός:** Κατά τη φάση σχεδιασμού, η ομάδα μας συγκεντρώθηκε για να προσδιορίσει τις απαιτήσεις του έργου και να δημιουργήσει το αρχικό UML διάγραμμα κλάσεων. Κάθε μέλος ανέλαβε συγκεκριμένες αρμοδιότητες για τον σχεδιασμό της εφαρμογής.
2. **Ανάπτυξη:** Το έργο χωρίστηκε σε μικρές ενότητες (sprints), με κάθε sprint να διαρκεί δύο εβδομάδες. Κατά τη διάρκεια αυτών των sprint:
 - Ο Θεοδόσης ανέπτυξε τον κεντρικό κώδικα και διασφάλισε την υλοποίηση των βασικών λειτουργιών.
 - Ο Φοίβος ήταν υπεύθυνος για την προετοιμασία και τον καθαρισμό των δεδομένων, εξασφαλίζοντας την ακρίβεια και την ποιότητα των χρησιμοποιούμενων δεδομένων.
 - Ο Κωνσταντίνος δημιούργησε διαγράμματα και γραφήματα για να βελτιώσει την κατανόηση των αποτελεσμάτων της ανάλυσης.
3. **Δοκιμές:** Στο τέλος κάθε sprint, η ομάδα διεξήγαγε δοκιμές για να διασφαλίσει τη σωστή λειτουργία της εφαρμογής. Αυτές οι δοκιμές περιελάμβαναν unit tests και ολοκληρωμένες δοκιμές της εφαρμογής.
4. **Παράδοση:** Στο τέλος κάθε sprint, παραδόθηκε μια ενημερωμένη έκδοση της εφαρμογής. Το GitHub χρησιμοποιήθηκε για τη διαχείριση του κώδικα και τη συνεργασία της ομάδας.

Η χρήση του Agile μας επέτρεψε να ανταποκριθούμε γρήγορα στις αλλαγές των απαιτήσεων και να βελτιώσουμε συνεχώς την εφαρμογή μας μέσω επαναληπτικών κύκλων ανάπτυξης.

8 Σύνδεσμοι

- <https://github.com/yaGmomoY/MessieApp>
- <https://www.overleaf.com/read/kqpmxswdxbygcd087e>



Εφαρμογή Ανάλυσης Δεδομένων

Επιλέξτε ένα αρχείο CSV ή Excel



Drag and drop file here


Limit 200MB per file • CSV, XLSX

Browse files


Figure 2: Screenshot λειτουργίας της Εφαρμογής

Εφαρμογή Ανάλυσης Δεδομένων

Επιλέξτε ένα αρχείο CSV ή Excel

 Drag and drop file here
Limit 200MB per file • CSV, XLSX

Browse files

 Παράδειγμα.csv 336.0B

✕

Data Preview:

	feature1	feature2	feature3	label
0	1	2.1	3.3	A
1	1.2	2	3.1	A
2	1.4	2.2	3	B
3	1.5	2.3	2.9	B
4	1.1	2.4	3.5	A

Data Info:

None

Label Column: label

Label is categorical

Processed Data:

	feature1	feature2	feature3	label
0	1	2.1	3.3	0

Figure 3: Screenshot λειτουργίας της Εφαρμογής

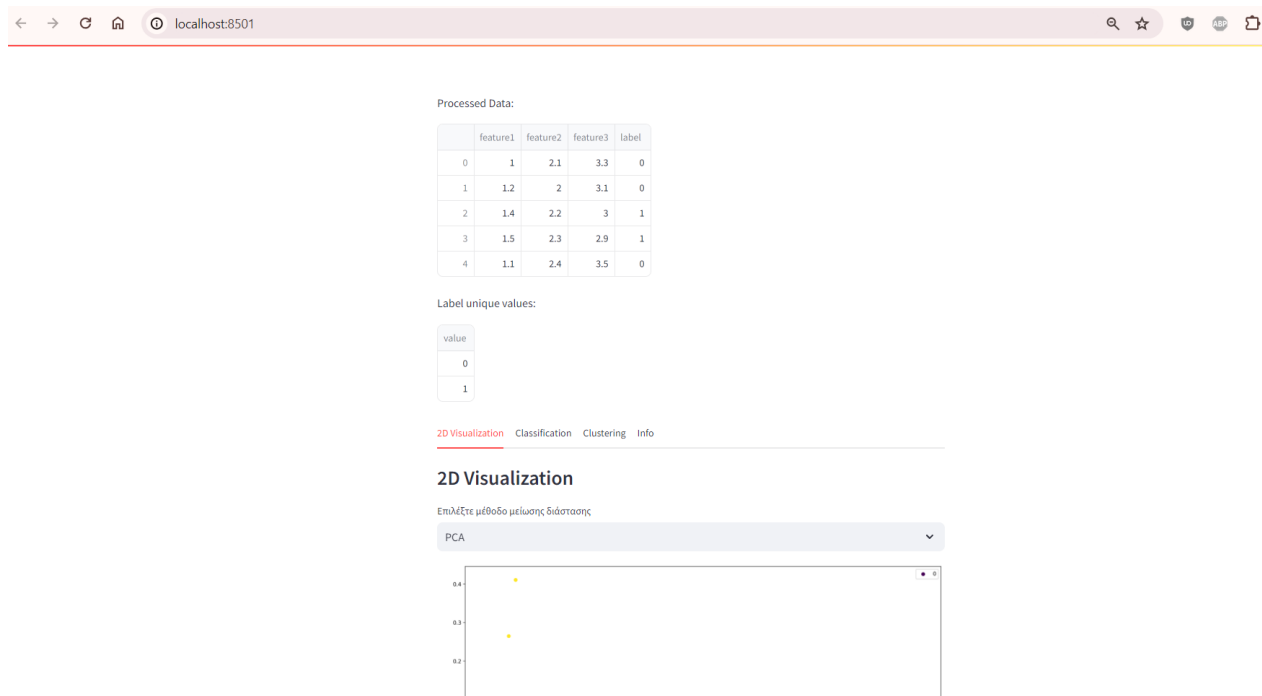


Figure 4: Screenshot λειτουργίας της Εφαρμογής

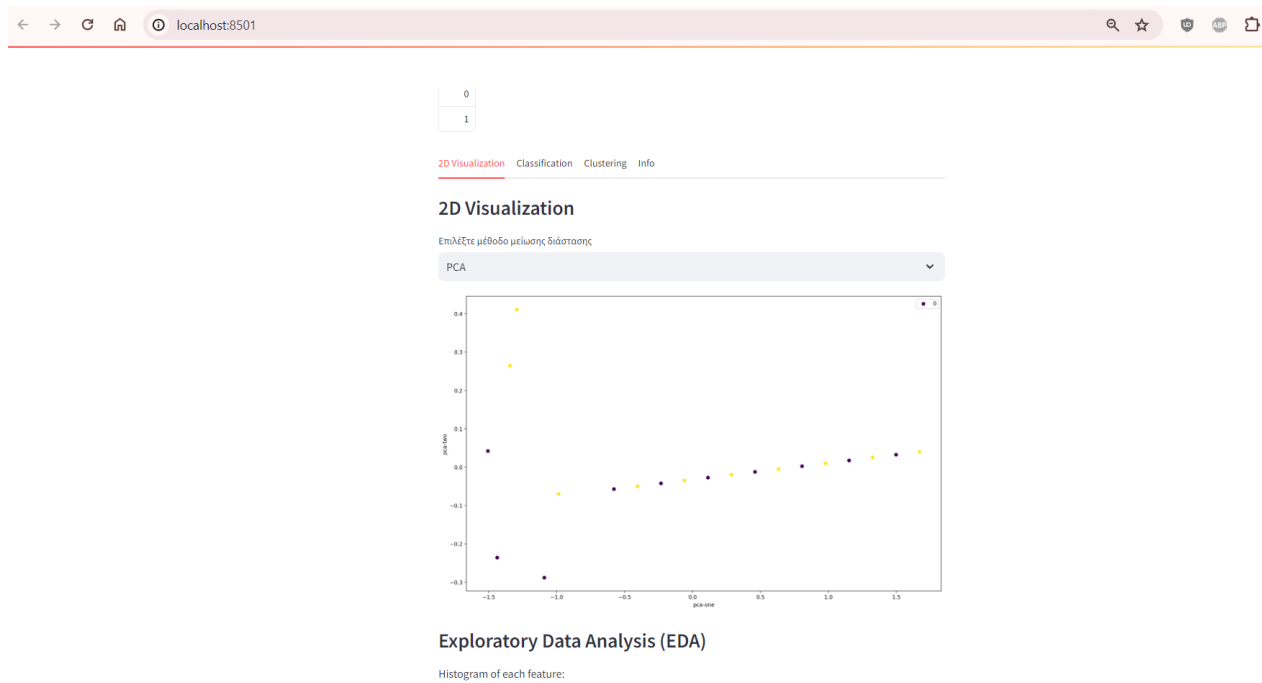


Figure 5: Screenshot λειτουργίας της Εφαρμογής

Exploratory Data Analysis (EDA)

Histogram of each feature:

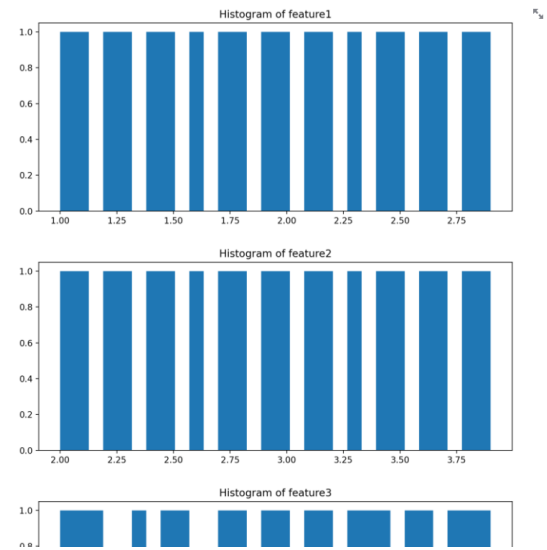


Figure 6: Screenshot λειτουργίας της Εφαρμογής

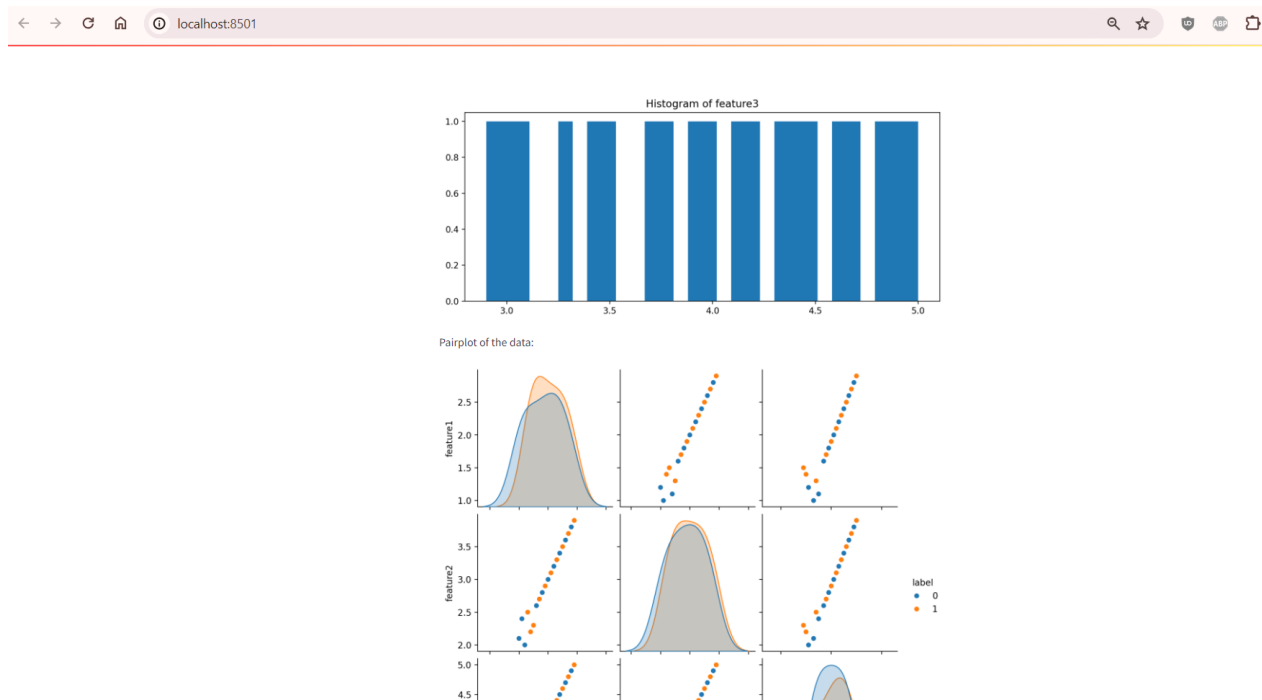


Figure 7: Screenshot λειτουργίας της Εφαρμογής

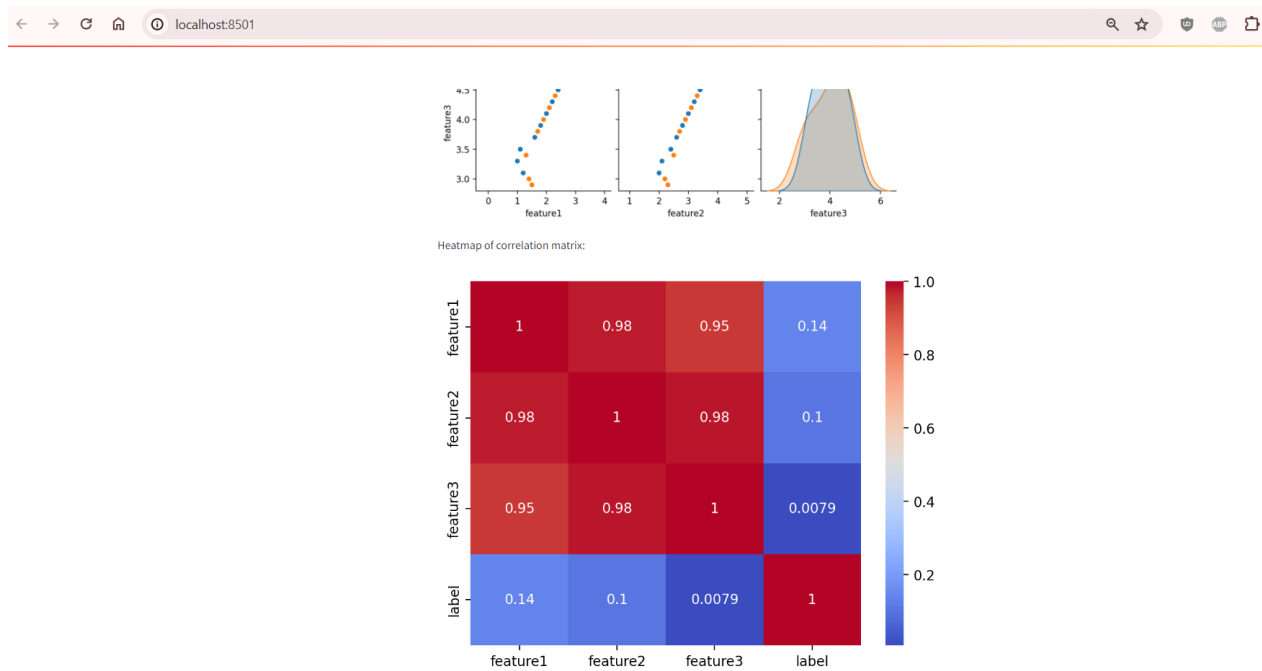


Figure 8: Screenshot λειτουργίας της Εφαρμογής

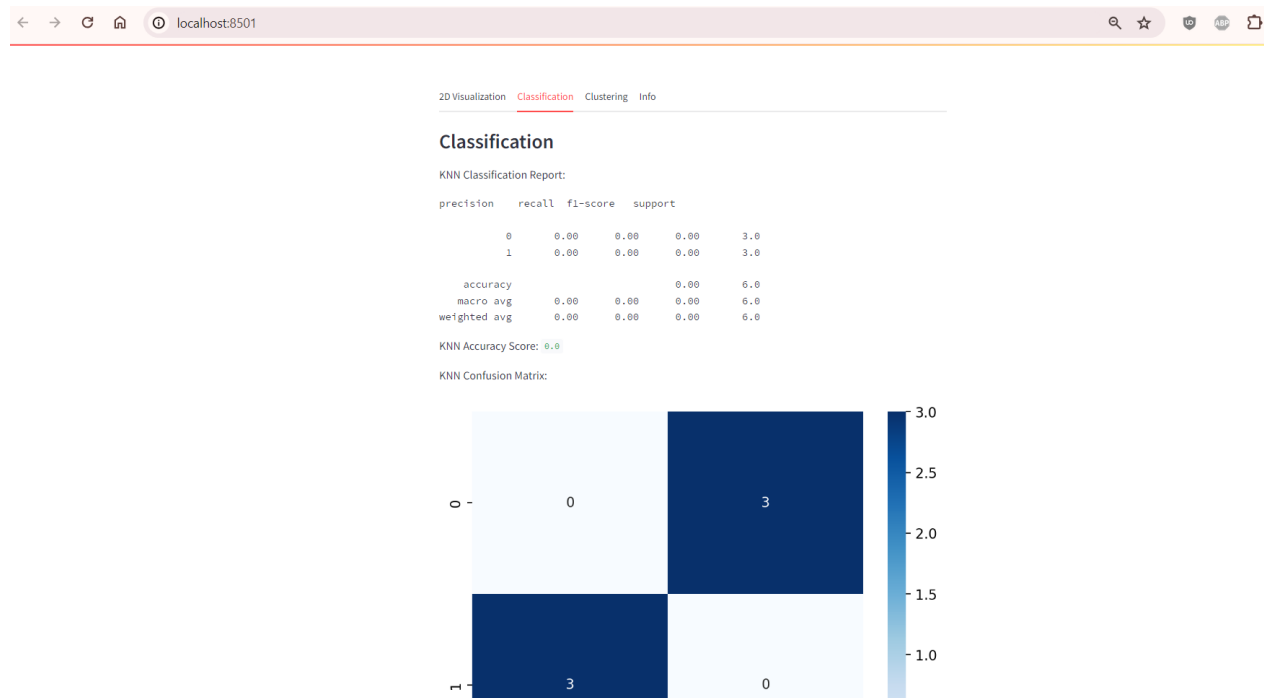


Figure 9: Screenshot λειτουργίας της Εφαρμογής



Figure 10: Screenshot λειτουργίας της Εφαρμογής

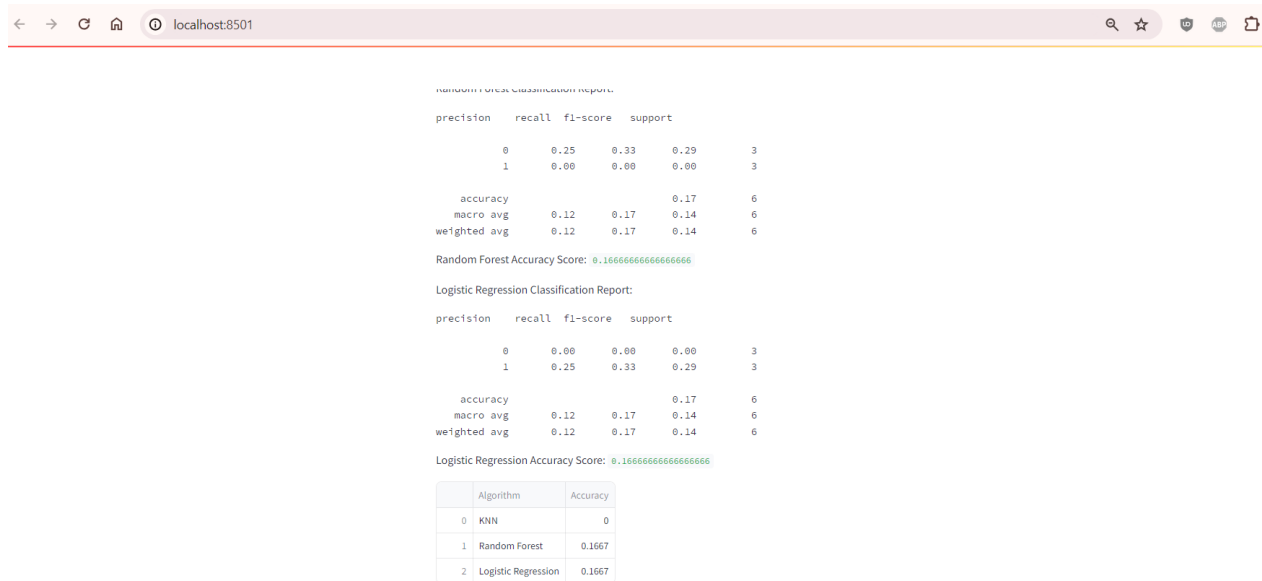


Figure 11: Screenshot λειτουργίας της Εφαρμογής

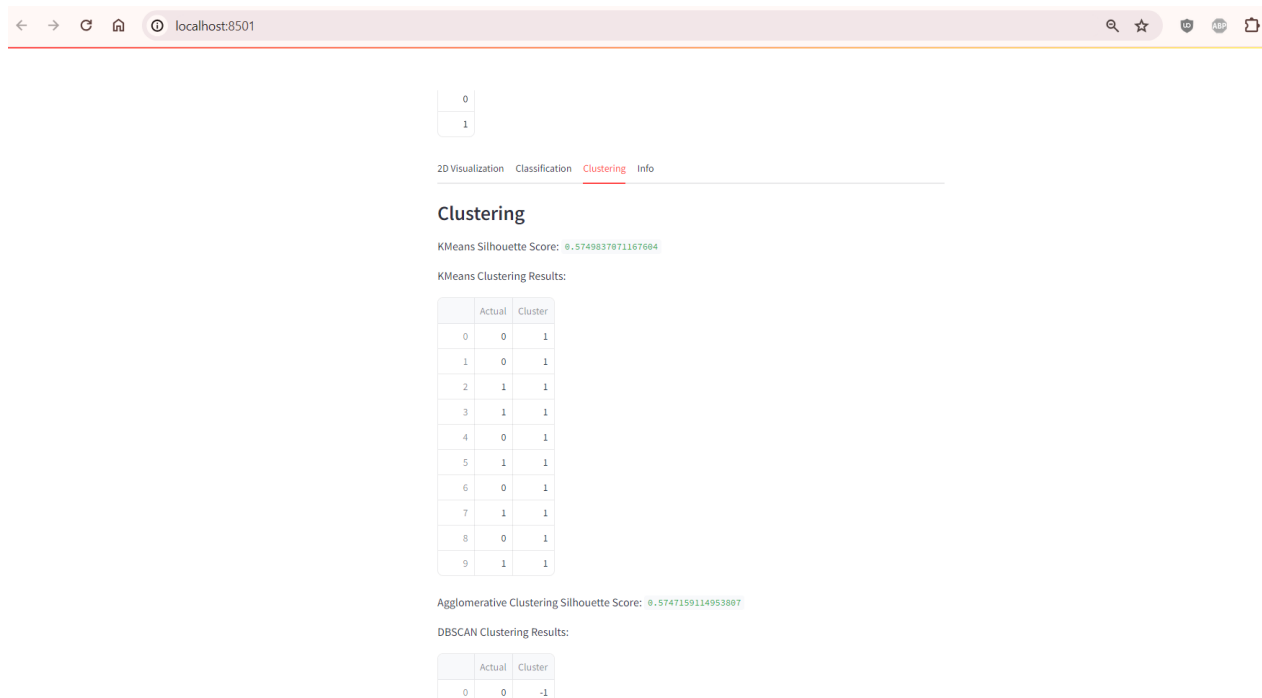


Figure 12: Screenshot λειτουργίας της Εφαρμογής

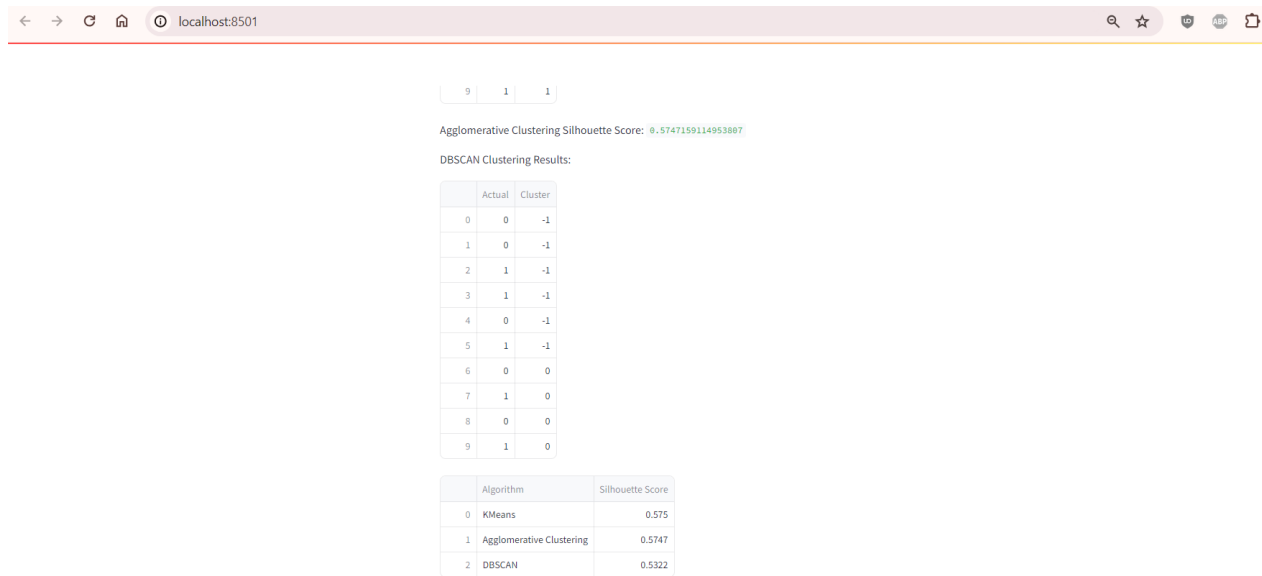


Figure 13: Screenshot λειτουργίας της Εφαρμογής

Info

Αυτή η εφαρμογή αναπτύχθηκε από την ομάδα μας για το μάθημα Τεχνολογία Λογισμικού.

Η εφαρμογή προσφέρει εργαλεία ανάλυσης δεδομένων, επιτρέπει τη φόρτωση, την οπτικοποίηση, την κατηγοριοποίηση και την ομαδοποίηση δεδομένων καθώς και την παρουσίαση των αποτελεσμάτων.

Λειτουργία της Εφαρμογής

1. Φόρτωση Δεδομένων: Επιλογή αρχείου CSV ή Excel
2. Οπτικοποίηση: 2D διαγράμματα με PCA ή t-SNE
3. EDA: Ιστογράμματα και pairplots για κατανόηση δεδομένων
4. Κατηγοριοποίηση: Αλγόριθμοι KNN και Random Forest
5. Ομαδοποίηση: Αλγόριθμοι K-Means και Agglomerative Clustering

Χρησιμοποιούμε τεχνολογίες Python, Streamlit, και scikit-learn.

Ομάδα Ανάπτυξης:

Θεοδότης Καραγεωργίου (AM: Inf2021076): Ανάπτυξη Κώδικα

Κωνσταντίνος Λιάβας (AM: Inf2021121): Ανάλυση Δεδομένων

Φοίβος Ελευθερίου (AM: Inf2021049): Οπτικοποίηση Δεδομένων

Figure 14: Screenshot λειτουργίας της Εφαρμογής